Department of Business Administration

UZH Business Working Paper Series

Working Paper No. 378

Rankings upon rankings – and no end in sight

Discussion of "Quantitative and Qualitative Rankings of Scholars" by Rost and Frey

Uschi Backes-Gellner

January 2011

Published as: "Rankings upon Rankings - and no End in Sight - Discussion of "Quantitative and Qualitative Rankings of Scholars" by Rost and Frey." *Schmalenbach Business Review*, 63(2011)1: 99-108. By Uschi Backes-Gellner.

University of Zurich, Plattenstrasse 14, CH-8053 Zurich,
http://www.business.uzh.ch/forschung/wps.html

UZH Business Working Paper Series

Contact Details

Uschi Backes-Gellner

University of Zurich

Department of Business Adminstration

Uschi Backes-Gellner

Rankings upon rankings – and no end in sight

Discussion of

Katja Rost and Bruno S. Frey: Quantitative and Qualitative Rankings of Scholars

Katja Rost and Bruno S. Frey address an important topic. They compare two kinds of rankings, a conventional allegedly „quantitative“ publication ranking, and a ranking based on the membership on editorial boards of academic journals, which they call „qualitative“. They find that the relation between the two rankings is not linear, but inversely u-shaped. Consequently, they argue that maximizing publication rankings may lead to a decline in research quality. Therefore, basing promotion decisions solely on publication rankings could be *counter*productive for science and hence should be avoided.

The importance of Rost/Frey's topic for the scientific community is indisputable, and the implied conclusion that promotion decisions should not be based solely on publication rankings is just as clear. But the importance and the conclusion are presumably the only points of consensus. The proposed new ranking, the empirical study, and the implicit assumptions on which their analysis is based are all controversial, and must be evaluated more critically. In what follows, I single out three points that I believe are to be seen critically: Are publication rankings rightfully called „quantitative", given that rankings based on the membership on editorial boards are supposedly „qualitative"? Is the proposed new indicator „number of board memberships" better for research progress than the criticized indicator „number of publications"? Would it be better to refrain from rankings in general? Other authors might emphasize further points or perhaps choose entirely different ones, but such differing views are only natural, given the function of, and heavy discussions on, rankings. Since there will never be definitive answers, our objective can only be open and rational discussion. Therefore, I would first of all like to thank the *Schmalenbach Business Review* for having encouraged this discussion.

Are publication rankings even rightfully called „quantitative". given that rankings based on the membership on editorial boards are supposedly „qualitative"?

The title of the essay suggests a comparison between *quantitative* and *qualitative* rankings.[1] Therefore, we must first ask whether the new rank-

[1] Rost/Frey do concede that „...quantitative and qualitative rankings are not strictly separable, since both contain elements of the other. The distinction is made solely for clarity ..." (S. 4). However, throughout the rest of the text it does not prevent them from treating publication indicators as if they measured „quantity", and as if number of board memberships measured „quality".

ing proposed by Rost/Frey, which are based on the number of memberships on editorial boards, can rightfully be described as *„qualitative“,* given that rankings based on the (weighted) number of publications are described as „quantitative“. Is this even a true comparison of „quantitative“ and „qualitative“ rankings? The answer is quite clearly “no“.

The act of counting alone suggests that both rankings are in fact *quantitative*. However each of the two rankings also indisputably contains qualitative aspects. In both rankings, the indicators are „quantitative“ in nature, but if they are well constructed, they also signify “quality”. Publication indicators always reflect quality if the number of publications is weighted based on journal quality (which should be the standard nowadays). Obviously, this is not a perfect way to depict quality. But then again, are indicators ever really perfect? The “new” indicator, which Rost/Frey call „qualitative“, is also simply a number. The “number” of memberships on editorial boards is clearly a quantitative indicator as well, and it is limited to measuring a single dimension. At the same time, the authors charge the allegedly „quantitative“ publication rankings with not adequately considering the qualitative aspects of research activity. This accusation is certainly not new, and neither is it completely wrong, but it applies in just the same way to Rost/Frey’s so-called „qualitative“ ranking. The number of board memberships is also a single dimension, and carries only limited quality information. In the end, nothing is perfect. Indicators always depict certain dimensions better and others worse. The question of which of these dimensions are most important is still unanswered. Unfortunately, Rost/Frey do not provide an answer.

To summarize, I can state at this point that the fundamental difference between „quantitative“ and „qualitative“ does not exist in the way that the title of the paper of Rost/Frey suggests. Both indicators are simultaneously quantitative and qualitative, and both have their own specific ad-

vantages and disadvantages. Consequently, the relevant research question is not whether an indicator can depict all performance dimensions perfectly, but whether certain indicators, given their unique shortcomings, are more or less useful for the scientific community. In other words, an evaluation of alternative indicators should provide arguments to decide which indicator „causes less harm“ or „has more benefits“ to the scientific community.

2. Is the proposed new indicator, „number of board memberships“, better for the scientific community than the old indicator „number of publications“?

The notions of what purpose a ranking should or should not serve and what, exactly, a scientific community should accomplish are insurmountably different. Thus, there cannot be a definitive, conclusive answer to this question. But we can at least use some objective and generally applicable criteria for determining the quality of any performance indicator to shed some light on the problem, namely we analyze an indicator's scope for manipulation, its multitasking problems, and its potential for systematic distortions of results.[2] These criteria should be fulfilled in every instance because otherwise, applying such performance indicators will inevitably cause biased or adverse incentives, which in turn would not be favourable to the scientific community as a whole.[3]

[2] Rost/Frey mention precisely the same criteria in part of their analysis, but they do not use them for a detailed comparison of the two indicators.

[3] For a discussion of the relative advantages and disadvantages of various performance indicators for professors, see also Pull (2009).

Scope for Manipulation

First, I compare the two indicators regarding the ease with which they can be manipulated – i.e., the *scope for manipulation* – and the resulting distortion of incentives.[4] A quick comparison of publication indicators and board membership indicators concerning the „scope for manipulation" raises doubts as to whether an indicator based on board memberships is a better indicator in this dimension. The question is, how easy is it to just raise the number of (weighted) publications without any, or with only very low cost, and how easy is it to just raise the number of board memberships? Is just increasing the indicator cheaper for publications or for board memberships? Of course, a comprehensive analysis is not possible within the scope of this discussion paper. But I can provide some preliminary analyses to indicate how far the two performance indicators may be affected by manipulations of the agent (in this case, the researcher). In this context it is important to note that this problem becomes especially important ex post, i.e., once an indicator is known to be used for rankings. Some indicators may have yielded valid performance analyses historically, but the validity of the results is negated once the indicator is known to be used for rankings.

So what is the possible extent of a cheap manipulation regarding the number of publications or the number of board memberships, respectively? I would argue that the possibility of manipulating the *number of publications* is limited, at least for high-quality journals. High quality journals will always have a shortage of printing space. Hence, if a journal does not want to lose its reputation for quality, it must decide very carefully what to accept for print in the limited number of pages they have. Ex-

[4] Rost/Frey argue that „Scholars would invariably find ways to „beat the system". This is undoubtedly true, but this makes it essential to ask which indicator allows the least possibility to „beat the system". It seems unlikely that this indicator is the number of editorial board memberships as will be shown in the following.

tending the number of journal pages (i.e., extending the number of pages per volume or increasing the number of annual issues of a journal is usually not feasible, since it causes very high production costs and is an indicator of decreasing quality in and of itself. Consequently, journal editors will avoid allocating some of their already limited pages to a low-quality paper simply as a favour to some researcher (whether the author is a part of an editor's network or not). So the costs of a manipulative publication decision (i.e., accepting a lower-quality paper for a favour) are extremely high because every bad paper takes up some of the already scarce space that could otherwise be used to print high quality papers. Thus the scope for manipulation is rather small, at least as long as publication indicators are quality weighted.

The situation is entirely different when we consider the scope for manipulation in the *number of memberships on editorial boards*. The list of board members can be extended almost at will without incurring any significant costs. For whatever reason a journal would decide to add another researcher to their editorial board, they would just add another name to the list of board members. In the worst case scenario, the names of the board members would not fit on the front page anymore and the list would have to be continued on the second or third cover page, which causes only (if any) negligible additional production costs. And the potential loss in reputation caused by adding some more names on the front page (or perhaps even the second or third page) is certainly substantially lower than the loss in reputation caused by a low-quality paper that is undeservedly accepted for publication. Hence, the direct and indirect costs of manipulating memberships on editorial boards are significantly lower.

However, it is important to note that Rost/Frey defend the quality of board memberships as an indicator by arguing that „A (chief) editor

wants to have scholars at hand who help him or her make the best possible decisions; a disreputable individual or person lacking expert knowledge is of little use“. But this argument is at least equally, and probably significantly more, valid for the decision on what papers to publish in a journal. What editor would want to publish a paper with obviously lacking quality? So if, as Rost/Frey suggest, the incentives of journal editors are correct for board memberships, then they should be at least as correct for publication decisions. In this respect, the two rankings are not systematically different, but fundamentally similar.

As an easy and by no means final way to test this argument empirically, I compare the number of published papers with the number of editorial board members over the last few decades in one of the top journals in the management field, the *Academy of Management Journal*. The data and the graphs shown in Figure 1 and 2 cover the time span between 1963 (the first available year in JStor) and 2010 (the last available edition).

Insert figure 1 and 2 around here

Figure 1 shows that the number of papers has barely increased over the last 40 years, and that it fluctuates between 24 and 54 (cf. Fig. 1)[5]. Figure 2 shows that the number of editorial board members increases continually, and towards the end, dramatically, beginning with nine and end-

[5] I calculated the number of papers per year by multiplying the number of papers in the first issue of the year with the number of issues per year. Thus, the graph already accounts for the increase of issues per year from four to six in 1993. Here the reader should also note that the increase in issues per year does not lead to an increase in the total number of papers per year.

ing with 159 (cf. Fig. 2).[6] This development is not necessarily due to manipulation, but at the very least it shows that there is great leeway in the number of board members.

If we also consider that according to Rost/Frey, the number of board memberships is first mentioned in 1991 as a performance indicator for researchers (Gibbons/Fish 1991), the results are exactly what we would expect according to incentive theory and what I argued in the section above: the number of board members should begin to rise in the mid-1990s, at the very latest, while the number of publications should remain more or less constant.

Of course one could now argue that it is only difficult to manipulate the number of publications in top journals, or at least in high-quality journals. Given the increasing number of journals of somewhat questionable quality that have recently begun to be published, one might say that journal pages – and, consequently, the possible number of publications – are not really scarce anymore. This increase would imply that publication indicators have also developed more scope for manipulation over time. This claim is undoubtedly true, but well-constructed publication rankings weigh the number of articles according to journal quality for exactly this reason. Some rankings today even go so far as to entirely exclude low-quality journals in their publication counts. Others raise the convexity of the weights towards the upper limit of the journal quality (as it has just been done in the Journal Ranking for Economics Journals of the German newspaper „Handelsblatt“).

[6] I simply counted the number of editorial board members in each first issue of the year. Since the page listing the editorial board members is missing in the JSTOR-Version of issue 1/1973, I used issue 1/1974 in this case.

Therefore, if we use the *scope for manipulation* as a way to determine the quality of a ranking indicator, the number of (weighted) publications is certainly a less problematic indicator than the number of board memberships.

Multitasking Problems

A second and important quality criterion for the two ranking indicators is the extent of multitasking problems, as already mentioned in Rost/Frey. Here, we must carefully compare the multitasking problems that evolve while publishing or while being a board member. So what are these other, multiple tasks, and is it reasonable to assume that the multitasking problems are higher for publication rankings as Rost/Frey suggest? According to Rost/Frey, the other tasks are „ (...) teaching, supporting young scholars, informing and advising the public, participating in university administration, or reading and reviewing the work of other scholars." Without conducting a comprehensive analysis, it is still obvious that Rost/Frey's argument on why publication indicators cause multitasking problems also holds for board memberships. [7] For instance, one of the problems of publication rankings that Rost/Frey identify is that "a publishing record is easy to measure, but performance on the other tasks is not". This statement is just as true for the *number of board memberships*, because board memberships are also easier to measure than performance indicators of other tasks. Furthermore, attempts to maximize board membership counts often depend heavily on networking activities and on the establishment of personal contacts. These activities certainly do not improve performance on the other tasks, such as teaching or supporting young scholars. On the contrary, the attempt to maximize publication counts is at least somewhat beneficial to other scientific tasks

[7] Pull (2009) gives a more detailed discussion of multitasking problems when measuring performance of professors.

(for example, teaching senior students; see Backes-Gellner/Zander (1989), who find a positive correlation between teaching higher-level courses and publication output for economics and management professors in Germany. At the very least maximizing publication counts leads to efforts to contribute to the current state of research. The same is hardly true for attempts to maximize personal networks.

Systematic distortion of results

In comparing the two indicators, we must also ask ourselves whether they systematically favour or discriminate against certain groups of scientists, or whether they are more likely to deliver unbiased results. Rost/Frey claim that „editorial boards clearly favour established scholars“, but that the same is true for „the number of publications and citations“. Unfortunately, the authors do not support this statement empirically. Personally, I would expect the opposite to be true. Younger generations have a fair chance of being published – even without a record – if their research is of outstanding quality. However, the membership on editorial boards clearly favours established scientists and leaves little chance for younger scholars, no matter how outstanding they may be.

This distortion is especially problematic in the context of producing research, because research progress and scientific innovations – like any other innovations – rely on new ideas and particularly on tapping the full potential of young talents. Although I also cannot present hard empirical evidence, I will use an analogy to empirically support my argument, at least provisionally. I use a phenomenon that caused similar distortion and which was hardly beneficial for the scientific system. My example is presentations at scientific conferences, which became accessible to young scholars only after the selection processes for accepting/inviting paper presentations were aligned with the acceptance criteria used for

papers in scientific journals. The presentations at the yearly conferences of the German Economic Association (Verein für Socialpolitik) are well documented and serve as an empirical example here. As late as the 1970s, there were only very few presentations. Selection decisions were made by a handful of established professors from a small pool of experts and colleagues who were all personally known to one another. Consequently, only very few researchers – all of them professors – ever had a chance to present their work. For instance, according to the annual conference proceedings printed in the book series of the German Economic Association, of the 19 speakers of the annual conference in 1970 and the 34 speakers in 1982, every single one was a senior researcher. Young researchers, if they were permitted to attend at all, were only allowed to listen to the presentations, but not to present. Only after the Association opened part of the conference in the 1990s and based its choice of papers on an anonymous selection process did young scholars have a chance to present their work to the broader scientific community. Accordingly, in later years young scholars began to catch up. Between 1997 and 2002 at least 58% of speakers were young scientists with a diploma or Ph.D., while professors only made up about 20% of speakers (Fabel/Lehmann/Warning (2003)).[8] Today, the yearly conference of the German Economic Association would be inconceivable without all the young scholars in the open part of the conference – and scientific progress undoubtedly and to a large extent depends on this open part of the conference. Other examples for this phenomenon include the annual meeting of the German Association of Business Economists (Pfingsttagung des Hochschullehrerverbandes für BWL) and countless other conferences. Scientific progress would certainly not be able to keep up with

[8] Bommer/Ursprung (1998) also show that the presentation activities are significantly positively correlated with publication intensity. In other words, both studies show how much the scientific output now depends on young scientists.

the pace it has today if all of these conferences were still biased in favor of established scholars and against young scholars, as they were in the past.

Interpretation of U-shaped Relationship Between Publications and Board Memberships

Finally, the question remains, what we can conclude from Rost/Frey's *empirical* results on the relation between publications and board memberships (given that their theoretical arguments favouring board memberships over publications are rather unconvincing as shown above). So what do we learn from editorial board memberships and from the inversely U-shaped relationship Rost/Frey find?

Rost/Frey claim: „Journals are indeed interested to appoint board members who show a constant publication history but they are *not interested* in having board members who show a very large number of publications". But the claim that journals are not interested in board members with very large number of publications is not supported by the available data. To be precise, the inversely U-shaped relationship only tells us that researchers who are outstandingly productive are less often members of editorial boards than researchers with a lower publication productivity. It is unclear whether the outstandingly productive researchers have not been asked to participate in a board; or whether they preferred not to participate; or whether there is a trade-off regarding time investment, and that those who do participate in boards become less productive afterwards. As a matter of fact, we do not know whether the empirical results originate from the demand side, or from the supply side. Hence, Rost/Frey's argument that journals are „ ... *not* interested in having board members who have a very large number of publications" is audacious.

However, the good news in this context of an inversely u-shaped relation is that researchers with the lowest publication productivity are obviously the least likely to participate in editorial boards. In the end, this result speaks to the quality of the editorial system as a whole (and in this case, it is irrelevant whether scientists with low publication activity were not asked or simply did not want to participate, most important is that they just did not participate).

What if Editorial Board Memberships Indeed Became a Common Performance Indicator?

Last but not least, when evaluating the inversely U-shaped relation with respect to its consequences for the scientific community, we must ask whether it would even be an efficient allocation of resources if outstandingly publishing researchers spent their time on more editorial boards, which could happen if there were new and effective incentives for editorial board memberships. If we assume that the best research talents are particularly scarce, and if we further assume that research production and research services are substitutes for one another as Rost/Frey implicitly assume in their multitasking argument, then it seems doubtful that a redistribution of efforts would be a more efficient allocation of resources. Incentives that stem from publication indicators encourage original research production, but at best board memberships only provide incentives for an intermediate output. Should this intermediate output become an end in and of itself due to misguided incentives, then membership on editorial boards could easily become a detriment to research productivity. In any case, before concluding from the inverted U-shape that there are distorting or inefficient incentives (in the upper part of the talent distribution) we need to further investigate the allocative problems in combination with the incentive effects at hand.

3. Is it better to refrain from rankings in general?

Here I also see a clear "No," because we (the business researchers community) would find ourselves back in the situation where we started more than 20 years ago: internationally backward, self-referential, and stuck in the past. This is the situation that first engendered the discussion about rankings in Germany, and when the first colloquia and papers on measuring research performance and its organizational determinants occurred. For instance, a group of (social) scientists at Schloss Reisensburg (cf. Fisch/Daniel 1986, Daniel/Fisch 1988) tried to determine why German research output had become less and less visible over the years and why the distribution of individual research output was so uneven across researchers. They analyzed if a lack of objective performance measurement led to reduced incentives and declining research output; whether research performance was even measurable in general; and how, based on international experience, a systematic measurement of research performance could (or could not) be implemented effectively. (For more information see also Backes-Gellner 1987, and Backes-Gellner/Moog 2003.) Given the situation at the time, the results of these discussions were clear and widely accepted: the lack of performance measurement did not automatically lead to great research success as those who criticized performance measurements sometimes seemed to suggest, but the existence of performance measurements did at least foster a minimum effort (of course without inhibiting maximum effort and great research). Instead, a lack of performance measurement led to virtually inexcusable differences in research output despite nearly identical input; it led to an unwillingness to accept external criticism; and it led to complacency instead of the desire to continually improve. All these problems have been extensively described and discussed, and need not be

repeated here. An obvious consequence of this discussion was an increased and shared understanding that research performance should, in principle, be measurable and that the profession as a whole would have to face up to the measurement of their performance – particularly as business economists, who spend a significant amount of their time on counting and evaluating the performance of others.

Both the business economics profession and the scientific system as a whole (especially in the German-speaking countries) have come a long way since - and based on measurable results it was the right way to go. Today, young scholars are included as a matter of course at conferences; young German-speaking scholars are major contributors of papers in leading international journals; and appointments to professorships are predominantly based on research output instead of on membership in „Old Boys Networks“, as was often the case in the past.[9] Today, researchers expect to have their performance measured just like everybody else, and recognize that they cannot completely insulate themselves without consequences. In a world in which everything and everybody is measured and ranked, it is probably a virtue in and of itself that researchers also accept to be evaluated.

Perhaps, as happens so often when reforms are implemented, the current situation with all kinds of rankings being applied everywhere is too extreme. But this extreme result cannot mean that we need only revert to the old system and hope for everything to improve. It can only mean that we might need to be more careful, that we might need to differentiate more precisely, that we might need to consider which rankings should serve which purpose, that we might need to weigh different rankings

[9] The increased population of female researchers among newly appointed professors in recent years, might also be an important side effect of the declining importance of the "old boys networks" and the increasing importance of objective criteria such as published research, conference presentations, teaching evaluations, etc.

against one another, and that we might need to let one ranking complement the other. In this vein, Rost/Frey's contribution is highly laudable because it introduces a new indicator – membership in editorial boards – that certainly also generates valid information. However, I hope that this discussion has shown that we should not see their indicator final word on rankings, nor should it even be perceived as „better" than publication rankings.

References:

Backes-Gellner, Uschi (1987) Ökonomie der Hochschulforschung - Organisationstheoretische Überlegungen und betriebswirtschaftliche Befunde. Wiesbaden: Gabler 1989.

Backes-Gellner, Uschi (1989) "Zum Verhältnis von Forschung und Lehre in sozialwissenschaftlichen Fachbereichen." *Ökonomie der Hochschule I*, (1989), Ed. Christof Helberger. Berlin, München: Duncker & Humblot, 51-76.

Backes-Gellner, Uschi; Petra Moog (2004) Ökonomie der Evaluation von Schulen und Hochschulen. Beiträge der Jahrestagung 2003 des Bildungsökonomischen Ausschusses. Berlin: Duncker & Humblot. Schriften des Vereins für Socialpolitik.

Backes-Gellner, Uschi; Eva Zanders (1989) Lehre und Forschung als Verbundproduktion - Data-Envelopment-Analysen und organisationsökonomische Interpretationen der Realität in wirtschaftswissenschaftlichen Fachbereichen. In: Zeitschrift für Betriebswirtschaft, 59(1989)3: 271-290.

Bommer, R., H.W. Ursprung (1998) Spieglein, Spieglein an der Wand. Eine publikationsanalytische Erfassung der Forschungsleistungen volkswirtschaftlicher Fachbereiche in Deutschland, Österreich und der Schweiz. In: Zeitschrift für Wirtschafts- und Sozialwissenschaften, 118(1998): 1-28.

Daniel, H.-D., R. Fisch (Eds.) 1988: Evaluation von Forschung, Konstanz: Universitätsverlag.

Fabel, Oliver, Erik Lehmann, Susanne Warning (2003) Vorträge als Qualitätsindikator: Empirische Evidenz der Jahrestagungen des Vereins für Socialpolitik. In: Backes-Gellner/Schmidtke (Hrsg.) Hochschulökonomie – Analysen interner Steuerungsprobleme und gesamtwirtschaftlicher Effekte. Berlin: Duncker&Humblot, S. 13-31.

Fisch, R.; H.-D. Daniel (Eds.) 1986: Messung und Förderung von Forschungsleistung. Konstanz: Universitätsverlag.

Frey, Bruno S. (2007) Evaluierungen, Evaluierungen, ... Evaluitis. Perspektiven der Wirtschaftspolitik 2008 8(3): 207-220.

Pull, Kerstin (2009) Risiken und Nebenwirkungen der Leistungsmessung von Professoren. DBW 69(2009)3: 311-313.

Fig. 1: Number of papers in AMJ per year

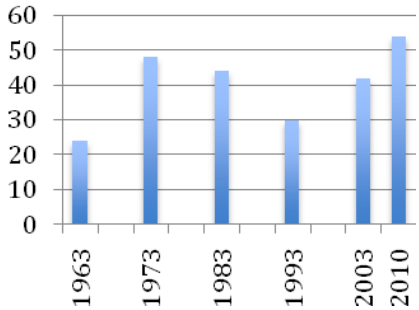

Fig. 2: Number of Editorial Board Members AMJ